Audio Declipping with Social Sparsity

Kai Siedenburg, Matthieu Kowalski, Monika Dörfler

AUDIO DECLIPPING WITH SOCIAL SPARSITY

Kai Siedenburg⋆*, Matthieu Kowalski*† *and Monika Dörfler*‡

⋆ CIRMMT, Schulich School of Music, McGill University Montreal, Canada
† CNRS-SUPELEC-Univ Paris-Sud, Gif-sur-Yvette, France
‡NuHAG, Faculty of Mathematics, University of Vienna, Austria

ABSTRACT

We consider the audio declipping problem by using iterative thresholding algorithms and the principle of *social sparsity*. This recently introduced approach features thresholding/shrinkage operators which allow to model dependencies between neighboring coefficients in expansions with time-frequency dictionaries. A new unconstrained convex formulation of the audio declipping problem is introduced. The chosen structured thresholding operators are the so called *windowed group-Lasso* and the *persistent empirical Wiener*. The usage of these operators significantly improves the quality of the reconstruction, compared to simple soft-thresholding. The resulting algorithm is fast, simple to implement, and it outperforms the state of the art in terms of signal to noise ratio.

Index Terms— Structured sparsity, Audio declipping, Iterative Shrinkage/Thresholding Algorithm

1. INTRODUCTION

1.1. Problem Statement

An important task in digital audio restoration is the recovery of missing or corrupted samples of a signal. Two important cases of this problem concern a) missing samples (or even full intervals of samples) and b) clipped audio. While the former mostly arises through errors of signal transmission, clipping denotes a situation in which a signal's amplitude exceeds a certain threshold and is truncated. For clipped signals, as opposed to the data loss case, at least the lost samples' correct sign values are known. Clipping is a common problem in digital audio systems whose maximum gain can be exceeded for many reasons. The resulting signal truncation leads to very unpleasant digital distortion.

Based on the assumption of sparse synthesis coefficients of the original signal, the declipping problem can be modeled by

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\boldsymbol{\alpha}\|_0 \quad \text{s.t.} \quad \|\mathbf{y}^r - \mathbf{M}^r\boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 \leq \epsilon \tag{1}$$

Here, $\boldsymbol{\alpha} \in \mathbb{C}^N$ denotes the synthesis coefficients and $\boldsymbol{\Phi} \in \mathbb{C}^{T\times N}$ the synthesis operator corresponding to the employed time-frequency dictionary. The vector $\mathbf{y}^r = \mathbf{M}^r\mathbf{y} \in \mathbb{R}^M$ denotes the reliable samples of the observed signal $\mathbf{y} \in \mathbb{R}^T$, that is, the unclipped samples in the clipping case or just the available samples in the case of data packet loss. Then, $\mathbf{M}^r \in \mathbb{R}^{M\times T}$ is a matrix comprised of those rows of the identity matrix that choose the entries of the reliable samples. Regarding the synthesis operator, Gabor frames (a.k.a. Short-Time Fourier-Transform, cf. [1, 2]) have proven to be well suited for the representation of audio signals, especially in the context of sparse decomposition. Therefore, $\boldsymbol{\Phi}$ will denote the matrix associated to a Gabor dictionary in the following.

In this paper, we specifically address the important problem of audio declipping. Here, an additional constraint can be added to (1): reconstructed samples must be greater (in absolute value) than the clipping threshold. In analogy to the definition of $\mathbf{M}^r$, let $\mathbf{M}^c \in \mathbb{R}^{(T-M)\times T}$ denote the matrix picking the clipped samples. Also, let $\boldsymbol{\theta}^{clip} \in \mathbb{R}^{(T-M)}$ be the vector of clipped samples, taking only the values $\pm\theta^{clip}$, in dependence on the sign of the true values in $\mathbf{y}$. Then, for declipping, the problem becomes

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\boldsymbol{\alpha}\|_0 \tag{2}$$
$$\text{s.t. } \|\mathbf{y}^r - \mathbf{M}^r\boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 \leq \epsilon \quad \text{and} \quad |\mathbf{M}^c\boldsymbol{\Phi}\boldsymbol{\alpha}| \geq |\boldsymbol{\theta}^{clip}|$$

1.2. Previous Work

Classical approaches to audio interpolation and declipping include autoregressive (AR) modeling [3], signal matching with bandwidth constraints [4], and Bayesian estimation [5]. A sparsity based formulation has been provided in [6], with (1) serving as the basic problem which was solved using orthogonal matching pursuit (OMP). The authors dubbed the method *audio inpainting* in reference to sparsity constrained image inpainting [7].

It is well known that the convex relaxation of (1) leads to the Lasso [8] or Basis Pursuit Denoising problem [9], yielding the following minimization problem:

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y}^r - \mathbf{M}^r\boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1 \,. \tag{3}$$

This convex non-smooth functional can be minimized by the popular iterative shrinkage/thresholding algorithm (ISTA) [10], also called forward-backward algorithm [11], or its accelerated version FISTA [12]. The most straight-forward approach for obtaining a convex relaxation of (2) with linear constraints is to consider (3) with the additional constraint to choose $\boldsymbol{\alpha}$, such that

$$|\mathbf{M}^c\boldsymbol{\Phi}\boldsymbol{\alpha}| \geq |\boldsymbol{\theta}^{clip}| \tag{4}$$

However, non-smooth convex problems with such constraints cannot be minimized directly by a forward-backward strategy. Indeed, the proximity operator of the ℓ_1 penalty with the linear constraint cannot be computed in closed-form, and one has to use an inner iteration inside the forward-backward algorithm to approximate it. Thus, a common approach is to use a Douglas-Rachford algorithm as inner loop, see e.g. [13], but this is usually accompanied by a very high computational burden.

KS is supported by a Harman Scholarship from the AES Educational Foundation. MK benefited from the support of the "FMJH Program Gaspard Monge in optimization and operation research", and from the support to this program from EDF. MD is funded by the Vienna Science and Technology Fund (WWTF) through project VRG12-009.

In the OMP based approach [6], the clipping constraint in (2), as well as an additional maximum constraint forcing the declipped signal below a maximum value, are satisfied by using a two-step strategy. First, OMP recovers the time-frequency support of the solution. Second, standard convex optimization solvers are applied on the obtained support. In comparison to traditional approaches, this approach leads to improvements in terms of signal to noise ratio, but has the shortcoming of being computationally very expensive, as it eventually employs high dimensional convex optimization.

An alternative iterative hard-thresholding formulation for declipping was recently described in [14]. Here, the iteration consists of a gradient descent step followed by hard-thresholding and thus bears similarity to iterative approaches as (F)ISTA. The authors reported improvements over the constrained OMP method [6], although their evaluation was based on a rather small set of audio examples. While the algorithm to be presented in the following also makes use of an ISTA-type iteration and shares some of the properties of the hard-thresholding approach, it was developed independently, cf. [15].

Besides the sparsity principle used in (1), many authors have investigated various *structured* sparsity approaches. This includes perceptually-informed compressed sensing [16] and Group-Lasso techniques [17] which allow to define grouping of coefficients to be jointly processed. Various extension to groups including overlaps have been proposed, such as the Latent-Group-Lasso [18, 19]. More specifically in the context of audio processing, Group-Lasso with overlap has been studied in [20]. The main drawback of these approaches is the high computational load required to solve the suitable functionals. The methods of *social sparsity* proposed in the current contribution avoid this practical drawback, cf. [21] for more detailed theoretical background.

This article features two main contributions. First, we propose an *unconstrained* convex relaxation of (2) which yields a solution with desired declipping behavior: the reconstructed samples' absolute values are above the clipping threshold. This unconstrained formulation allows to employ ISTA-type algorithms. Second, we explore the benefits of the recently introduced concept of social sparsity [21], in order to take into account temporal dependencies between Gabor synthesis coefficients. The remainder of the paper is organized as follows. Section 2 introduces the unconstrained convex formulation for the declipping problem and derives the associated ISTA algorithm. Section 3 is a brief recap of social sparsity and associated structured shrinkage operators to be embedded in ISTA in practice. Section 4 presents numerical experiments on the declipping problem and compares the approach with the algorithms presented in [6] and [14].

2. AN UNCONSTRAINED CONVEX FORMULATION

In this contribution we propose to relax the constraints (4) by means of a *squared hinge* function. This is a well-known function in classification, see e.g. [22], defined as follows:

$$h^2 \ : \ \mathbb{R} \longrightarrow \mathbb{R}_+ \quad z \mapsto h^2(z) = \begin{cases} z^2 & \text{if } z < 0 \\ 0 & \text{if } z \geq 0 \end{cases}$$

By application to $z = x - \theta^{clip}$, for known clipping values $\theta^{clip} > 0$, the squared hinge sets x "free" if $|x| \geq \theta^{clip}$, and penalizes otherwise. Using the notation

$$[\boldsymbol{\theta}^{clip} - \mathbf{x}]_+^2 = \sum_{k:\theta_k^{clip}>0} h^2(x_k - \theta_k^{clip}) + \sum_{k:\theta_k^{clip}<0} h^2(\theta_k^{clip} - x_k)$$

we thus introduce the following *unconstrained* convex optimization problem

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y}^r - \mathbf{M}^r\boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \frac{1}{2}[\boldsymbol{\theta}^{clip} - \mathbf{M}^c\boldsymbol{\Phi}\boldsymbol{\alpha}]_+^2 + \lambda\|\boldsymbol{\alpha}\|_1 \quad (5)$$

Since the squared hinge is differentiable with Lipschitz-continuous gradient, cf. [22], so is

$$\boldsymbol{\alpha} \mapsto \frac{1}{2}\|\mathbf{y}^r - \mathbf{M}^r\boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \frac{1}{2}[\boldsymbol{\theta}^{clip} - \mathbf{M}^c\boldsymbol{\Phi}\boldsymbol{\alpha}]_+^2 \quad (6)$$

and any algorithm from the ISTA family can be applied to solve (5), cf. [11, 12].

3. SOCIAL SPARSITY AND THE PROPOSED ALGORITHM

For approximating a solution to the relaxed problem (5), our work explores the application of social sparsity operators [21]. Social sparsity allows to incorporate a priori knowledge about signal classes and artifacts. Given the structure of the declipping problem it is natural to take temporal correlation into account: signal components such as harmonics which extend over time induce temporally persistent coefficients. On the other hand, isolated high-energy coefficients or temporally localized spread of energy over frequency may be attributed to the corruption of the signal and should therefore be discarded in the reconstruction process. By extending the usual soft thresholding, corresponding to the ℓ^1 constraint as used in (5), it becomes possible to exploit the persistence-properties of signal components through time-frequency neighborhood systems [21]. Note that these generalized operators do not directly correspond to the minimization of a convex functional any more.

Denote by $\mathcal{N}(t)$ the set of indices forming the neighborhood of the index t for time-frequency coefficients $\boldsymbol{\alpha} = \{\alpha_{tf}\}$ and set $(x)^+ = \max(x, 0)$. We can then restate the classic Lasso and its persistent variation, the so-called Windowed Group-Lasso (WGL) [23]:

- Lasso : $\tilde{\alpha}_{tf} = \mathbb{S}_\lambda^L(\alpha_{tf}) = \alpha_{tf}\left(1 - \frac{\lambda}{|\alpha_{tf}|}\right)^+$
- WGL : $\tilde{\alpha}_{tf} = \mathbb{S}_\lambda^{WGL}(\alpha_{tf}) = \alpha_{tf}\left(1 - \frac{\lambda}{\sqrt{\sum_{t'\in\mathcal{N}(t)}|\alpha_{t'f}|^2}}\right)^+$

The application of the shrinkage operators associated to Lasso and WGL, respectively, typically leads to a loss of energy in the estimated signal. Thus, in practice, it is common to first use the Lasso in order to select the relevant time-frequency atoms, and to perform a least-square estimation of the signal w.r.t. the selected atoms [24] of the employed dictionary in a second step. Another strategy, not requiring the least-squares step, is to design thresholding operators which preserve the energy in the big coefficients, and which can also be used inside ISTA [25]. The most well known is the Empirical Wiener (EW) operator [26], also known as nonnegative garrote shrinkage [25]. The EW operator features an altered exponentiation of the coefficient energy while having the same support as the Lasso. Such an exponentiation can also be used on the WGL operator, yielding the persistent EW (PEW) [27]. These operators read

- EW : $\tilde{\alpha}_{tf} = \mathbb{S}_\lambda^{EW}(\alpha_{tf}) = \alpha_{tf}\left(1 - \frac{\lambda^2}{|\alpha_{tf}|^2}\right)^+$
- PEW : $\tilde{\alpha}_{tf} = \mathbb{S}_\lambda^{PEW}(\alpha_{tf}) = \alpha_{tf}\left(1 - \frac{\lambda^2}{\sum_{t'\in\mathcal{N}(t)}|\alpha_{t'f}|^2}\right)^+$

These generalized thresholding operators, denoted by $\mathbb{S}_\lambda$ for a threshold λ, are subsequently used in the ISTA-framework. Algorithm 1 lays out the details and is called *relaxed forward backward*. Here, γ denotes the coefficient of relaxation. For the Lasso, the convergence of the iterates $\alpha^{(k)}$ towards a minimizer of (5) can be proven for $-1 < \gamma < 1/2$ and the convergence of the value of the minimization functional towards its minimum for $1/2 \leq \gamma < 1$, see [11].

Algorithm 1: relaxed version of ISTA

Initialization: $\boldsymbol{\alpha}^{(0)} \in \mathbb{C}^N$, $\mathbf{z}^0 = \boldsymbol{\alpha}^{(0)}$, $k = 1$, $\delta = \|\boldsymbol{\Phi}\boldsymbol{\Phi}^*\|$
repeat
$\mathbf{g}1 = -\boldsymbol{\Phi}^*\mathbf{M}^{r^T}(\mathbf{y}^r - \mathbf{M}^r\boldsymbol{\Phi}\mathbf{z}^{(k-1)})$;
$\mathbf{g}2 = -\boldsymbol{\Phi}^*\mathbf{M}^{c^T}[\boldsymbol{\theta}^{clip} - \mathbf{M}^c\boldsymbol{\Phi}\mathbf{z}^{(k-1)}]_+$;
$\boldsymbol{\alpha}^{(k)} = \mathbb{S}_{\lambda/\delta}\left(\mathbf{z}^{(k-1)} - \frac{1}{\delta}(\mathbf{g}1 + \mathbf{g}2)\right)$;
$\mathbf{z}^{(k)} = \boldsymbol{\alpha}^{(k)} + \gamma(\boldsymbol{\alpha}^{(k)} - \boldsymbol{\alpha}^{(k-1)})$;
$k = k + 1$;
until *convergence*;

4. EXPERIMENTS

4.1. Setup

We choose a tight Gabor frame[1] as time-frequency dictionary $\boldsymbol{\Phi}$ in Algorithm 1. The frame is based on a Hann window of 1024 samples length (about 64 ms at 16 kHz audio sampling frequency) and a time-shift of 256 samples. Note that the corresponding analysis operator $\boldsymbol{\Phi}^*$ is also known as the *Sliding Window* or *Short-Time* Fourier Transform. Concerning the relaxation coefficient γ in Algorithm 1, we observed empirically that the choice of $\gamma = 0.9$ leads to an algorithm which is faster than FISTA and less prone to numerical errors. The parameter γ will thus be held constant in the following. We evaluate declipping performance using the measure of SNR_m which measures estimation quality on the clipped, i.e. missing values only. For a clipped signal y and its estimation $\hat{y}$, it is computed as $\mathrm{SNR}_m(y, \hat{y}) = 20 \log \frac{\|\mathbf{M}^c y\|}{\|\mathbf{M}^c(y-\hat{y})\|}$.

For the subsequently described experiments, we used audio data provided by `http://small-project.eu/` and employed for the evaluation of the respective audio inpainting toolbox [6]. Specifically, our evaluations are based on the toolbox's speech and music data sets, sampled at 16kHz, containing 10 different signals, each of 5 seconds duration. All signals were range-normalized, in order to have sample values lower than 1, and consecutively clipped at levels $0.1, 0.2, \ldots, 0.9$.

The operator abbreviated by OMP in the following refers to the min-max-constrained orthogonal matching pursuit, the best performing operator in [6]. We also include results from [14]; however, in order to avoid the introduction of bias due to a different transform, in our evaluation we use a Gabor transform with the above mentioned settings instead of the discrete Cosine transform employed within the consistent iterative hard-thresholding algorithm (HT) proposed in [14]. Here, the corresponding algorithm is run on the entire signal instead of a windowed version, and we use the strategy exposed in section 4.2 to decrease the thresholds λ. This allows for the same basic setup for all the algorithms. Sound examples featuring all mentioned algorithms can be found under `http://homepage.univie.ac.at/monika.doerfler/StrucAudio.html`.

[1] A frame is tight, if $\boldsymbol{\Phi}\boldsymbol{\Phi}^* = c \cdot \boldsymbol{I}$, for some positive constant c and $\boldsymbol{I}$ the identity operator [1].

4.2. Basic Properties

Regarding the choice of hyperparameter λ, we here use the classical "warm start" strategy [28], starting with a relatively large λ which decreases in every K^{th} step of the iteration (here $K = 500$). This method essentially simulates the choice of a small λ but circumvents the slow convergence of ISTA that such a choice usually implies. Note that since we do not have to deal with additive noise in the declipping scenario, low levels of the hyperparameter λ are fully appropriate. Fig. 1 exemplifies the corresponding evolution of the algorithm's SNR versus the hyperparameter λ, which assumes 10 values logarithmically spaced from $\lambda = 10^{-1}$ to $\lambda = 10^{-4}$. This example is based on the first music signal of the evaluation set, clipped at $\boldsymbol{\theta}^{clip} = 0.2$. The figure clearly shows the benefit of the warm start strategy. Especially PEW gains SNR rapidly during the first 500 seconds of runtime (executed on a standard consumer laptop). During that time, each decrease of λ seems to boost convergence significantly. Note that PEW achieves good performance in a computation time far below that of OMP.

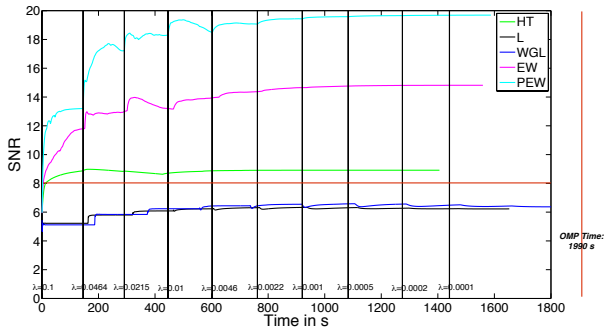

Fig. 1. Improvement of SNR as a function of time for a music signal at clip level $\boldsymbol{\theta}^{clip} = 0.2$. Algorithm 1 "warm-starts" with a large value of λ which is decreased every 500 iterations. Black lines indicates updates of λ. The horizontal red line indicates the SNR reached by OMP. The vertical red line indicates the cpu time taken by OMP.

A qualitative example of the declipping results is displayed in Fig. 2. Again, music signal no. 1 with clipping level $\boldsymbol{\theta}^{clip} = 0.2$ is shown, displaying estimations from all aforementioned operators. Neighborhoods for WGL and PEW extend 7 coefficients in time. On this time scale, it is obvious that (P)EW , HT and OMP operators yield much better estimations than the (WG)L: they reliably respect the clipping constraint (as do the (WG)L), and often resemble the original signal profile beyond the clip level, although OMP seems to yield too much high frequency oscillation, while HT sometimes overshoots the original amplitude values. The main differences between unclipped original and (P)EW-declipped version seem to be due to the shape of the largest amplitude values far beyond the clipping level. The (WG)L family, on the contrary, far more resembles the clipped signal than the original. In this case, the inclusion of neighborhood persistence does not even seem to yield a different solution, as both estimates behave almost identically.

4.3. Choice of the neighborhood

The choice of the neighborhoods used in the persistent thresholding operators WGL and PEW is both significant and delicate. A good choice of both size and, to a lesser extent, shape depends on the signal characteristics and the severity of corruption. Here, we evaluate neighborhoods which symmetrically extend in time and encompass

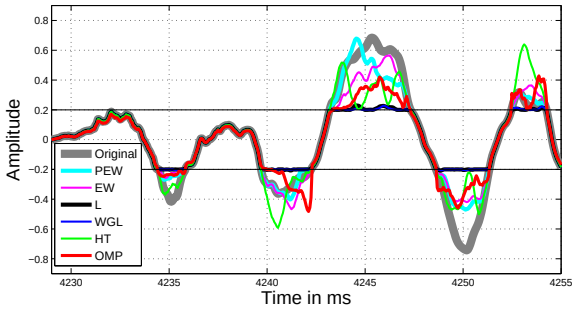

Fig. 2. Declipped music signal using different operators for clip level $\boldsymbol{\theta}^{clip} = 0.2$ using the Lasso, WGL, EW, PEW, HT, and OMP operators. Neighborhood size for WGL and PEW was 7.

3, 5, and 7 coefficients. Neighborhoods with 3 coefficients, for instance, would encompass the centre-coefficient itself plus one coefficient preceeding and one following in time. Note that the WGL and PEW with unit-neighborhood (with only one coefficient) coincide with the Lasso and EW operator, respectively. Fig. 3 depicts the average gain (over all clipping levels) of SNR_m obtained by using neighborhoods in conjunction with PEW and Algorithm 1, compared to the EW operator as a baseline, i.e. graphing SNR_m(PEW) $-$ SNR_m(EW). Here, applying neighborhoods improves reconstruction in about 50% of the cases, detoriation occurs otherwise. It is particularly visible that shorter temporal persistence of 3 coefficients brings more benefits than larger neighborhoods. For music signals, the usage of neighborhoods turns out to be favorable in almost all cases with best results for neighborhoods of size 5 or 7. We thus use neighborhoods of length 3 for speech signals and length 7 for music signals in the following comparisons.

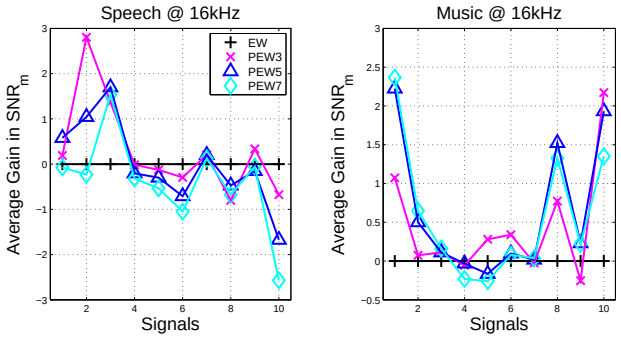

Fig. 3. Influence of the neighborhood on declipping performance. The 10 different speech (left) and music signals (right) are displayed on the x-axis. The average gain (over clipping levels $0.1, 0.2, \ldots, 0.9$) of SNR_{miss} with regard to the baseline set by the EW operator is displayed on the y-axis.

4.4. Evaluation on Speech and Music Signals

Fig. 4 presents systematic declipping results, depicting improvement of SNR_m (with respect to the baseline of the clipped signal) as a function of clipping level. Here, each point represents the mean over the ten different signals of the evaluation set.

Obviously, all operators improve SNR_m. However, Lasso and WGL seem to be least successful overall. This confirms the qualitative insights from Fig. 2: the better preservation of signal energy of the family of empirical Wiener-based operators yields more reliable

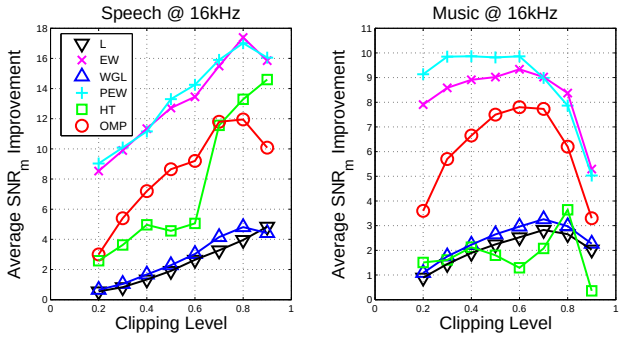

Fig. 4. Average SNR_{miss} for 10 speech (left) and music (right) signals over different clipping levels and operators. Neighborhoods extend 3 and 7 coefficients in time for speech and music signals, respectively.

estimates than the Lasso/WGL operators. No matter whether we consider WGL or PEW, however, the usage of neighborhoods still seems to improve performance for most clipping levels. Iterative hard-thresholding (HT) [14] has similar performance compared to L and WGL for music signals, but yields consistently better results on speech, where it is close to OMP [6]. Interestingly, for both speech and music signals, EW and PEW lead to improvements of up to 5 dB SNR_m compared to OMP. For low clipping values (i.e. massive signal deterioration) of music signals, the inclusion of neighborhood persistence seems to be particularly beneficial, yielding another 1 dB SNR_m improvement.

Except for the Lasso, the algorithms could in principle be sensitive to their initialization. We observed in practice that all operators, with HT being an exception, are very robust indeed and can be initialized by the clipped signal as a default value. However, HT is known to be very sensitive to initialization [29], which might partly explain the disappointing results obtained here.

Let us finally note that although no formal listening experiments were conducted for the lack of resources, we noticed that PEW-estimates feature remarkably few audible artifacts, even in cases of massive signal deterioration by low clipping thresholds. The approach thus not only performs well for improvement of signal to noise ratio, but seems to present a valuable tool for perceptual audio enhancement.

5. CONCLUSION

This paper studied the audio declipping problem by equipping the novel approach of social sparsity with a clipping constraint. We presented an algorithm which converges to a solution in the classical Lasso case and demonstrated empirically that thresholding operations beyond simple soft-thresholding lead to significant gain in declipping quality. Relevant improvements were achieved by taking into account alternative coefficient exponentiation (leading from L to EW) and neighborhood persistence (leading from Lasso to WGL and EW to PEW). In particular, the PEW operator yields significant improvements of SNR_m compared to two state-of-the-art methods, while the corresponding algorithm is still less time-consuming.

Future theoretical work will focus on the characterization of PEW operators in conjunctions with ISTA-type algorithms. Furthermore, we will apply the approach to the general audio interpolation (inpainting) problem, where exploiting neighborhood persistence in the time-frequency domain might be a valuable strategy for dealing with massive data loss such as time intervals of 10ms and more.

6. REFERENCES

[1] H. G. Feichtinger and T. Strohmer, "Introduction," in *Gabor Analysis and Algorithms Theory and Applications*, ser. Applied and Numerical Harmonic Analysis, H. G. Feichtinger and T. Strohmer, Eds., NuHAG. Birkhäuser Boston, 1998, pp. 1–31, 453–488.

[2] M. Dörfler, "Time-frequency Analysis for Music Signals. A Mathematical Approach," *Journal of New Music Research*, vol. 30, no. 1, pp. 3–12, 2001.

[3] A. Janssen, R. Veldhuis, and L. Vries, "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 34, no. 2, pp. 317–330, 1986.

[4] J. S. Abel and J. O. Smith III, "Restoring a clipped signal," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 1991, pp. 1745–1748.

[5] S. J. Godsill and P. J. Rayner, "A Bayesian approach to the restoration of degraded audio signals," *Speech and Audio Processing, IEEE Transactions on*, vol. 3, no. 4, pp. 267–278, 1995.

[6] A. Adler, V. Emiya, M. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 3, pp. 922–932, 2012.

[7] M. Elad, J.-L. Starck, D. L. Donoho, and P. Querre, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Journal on Applied and Computational Harmonic Analysis*, vol. 19, pp. 340–358, November 2005.

[8] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Serie B*, vol. 58, no. 1, pp. 267–288, 1996.

[9] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[10] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, August 2004.

[11] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York: Springer, 2011.

[12] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[13] C. Chaux, J.-C. Pesquet, and N. Pustelnik, "Nested iterative algorithms for convex constrained image recovery problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 730–762, 2009.

[14] S. Kitić, L. Jacques, N. Madhu, M. P. Hopwood, A. Spriet, and C. De Vleeschouwer, "Consistent iterative hard thresholding for signal declipping," in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 5939 – 5943.

[15] K. Siedenburg, M. Dörfler, and M. Kowalski, "Audio inpainting with social sparsity [abstract]," in *SPARS (Signal Processing with Adaptive Sparse Structured Representations), July 8-11, 2013, EPFL, Lausanne*, 2013.

[16] B. Defraene, N. Mansour, S. De Hertogh, T. van Waterschoot, M. Diehl, and M. Moonen, "Declipping of audio signals using perceptual compressed sensing," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2627 – 2637, 2013.

[17] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society Serie B*, vol. 68, no. 1, pp. 49–67, 2006.

[18] L. Jacob, G. Obozinski, and J.-P. Vert, "Group lasso with overlap and graph lasso," in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 433–440.

[19] G. Obozinski, L. Jacob, and J.-P. Vert, "Group lasso with overlaps: the latent group lasso approach." Tech. Rep., 2011.

[20] I. Bayram, "Mixed norms with overlapping groups as signal priors," in *Proceedings of the International Conference on Audio Speech and Signal Processing (ICASSP)*, May 2011, pp. 4036–4039.

[21] M. Kowalski, K. Siedenburg, and M. Dörfler, "Social sparsity! Neighborhood systems enrich structured shrinkage operators," *IEEE transactions on signal processing*, vol. 61, no. 10, pp. 2498–2511, 2013.

[22] M. Kowalski, M. Szafranski, and L. Ralaivola, "Multiple indefinite kernel learning with mixed norm regularization," in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 545–552.

[23] M. Kowalski and B. Torrésani, "Sparsity and persistence: mixed norms provide simple signals models with dependent coefficients," *Signal, Image and Video Processing*, vol. 3, no. 3, pp. 251–264, 2009.

[24] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, pp. 586–598, 2007.

[25] A. Antoniadis, "Wavelet methods in statistics: Some recent developments and their applications," *Statistics Surveys*, vol. 1, pp. 16–55, 2007.

[26] S. P. Ghael, A. M. Sayeed, and R. G. Baraniuk, "Improved wavelet denoising via empirical wiener filtering," in *Optical Science, Engineering and Instrumentation'97*. International Society for Optics and Photonics, 1997, pp. 389–399.

[27] K. Siedenburg and M. Dörfler, "Persistent time-frequency shrinkage for audio denoising," *Journal of the Audio Engineering Society (AES)*, vol. 61, no. 1, 2013.

[28] I. Loris, "On the performance of algorithms for the minimization of 1-penalized functionals," *Inverse Problems*, vol. 25, no. 3, pp. 35 008–35 023, 2009.

[29] M. E. D. T. Blumensath, "Iterative thresholding for sparse approximations," *The Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 629–654, 2008.